Poster No. 83

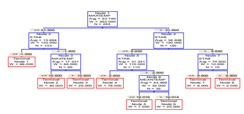
Application of the Categorical and Regression Tree (CART) Model (trademark Salford Systems, San Diego, Ca) to Understand the Relationship between PM and Meteorological Variables in the San Joaquin Valley

Richard J Hackney and Jeff Austin, California Air Resources Board, Planning and Technical Support Division, Sacramento, California

CART – what is it?

- CART-- statistical model that relates PM10 or PM2.5 concentrations and independent meteorological variables such as RH, temperature (min, max, mean), stability, visibility, and precipitation.
- The output is trees that show terminal nodes relating how much the meteorological variables contribute to PM concentrations (the nodes) as well as the relevancy of each meteorological variable.

Figure 1 -- Example CART Tree – Splitting the Met Variables



CART STUDY DOMAIN

CRPAQS Meteorology Sites for CART Analysis

CART Model Input Requirements

- Data, air quality and met consisted of:
  - Every 6th day sampling of PM10 and PM2.5
  - For four sites – Bakersfield, Fresno, Stockton, and Corcoran
  - 1988 to 2000 for PM10
  - 1999 to 2000 for PM2.5
  - Meteorology consisted of surface and Oakland aloft at the 4 stations above – most all surface type parameters, 850 mb temp, 500 mb height and stability (850Temp-closest minimum surface temp)

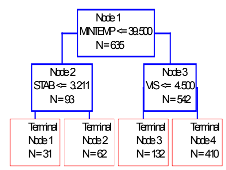
MODEL OPTIONS

- Regression tree, least squares option
- Rest of options model default options
- Run from 1988 to 2000 for PM10 and 1999 to 2000 for PM2.5, for all seasons, fall (September to November), and winter (December to February)
- Ran for four separate regions of the SJV – north (Stockton), central (Fresno), south (Bakersfield) and agricultural (Corcoran)
- Computer run times of a few minutes or less allowing many corrective runs to be performed daily

CART ANALYSIS USED FOR SAN JOAQUIN VALLEY PM10 SIP– ALL SEASONS

- Figure 2 -- tree for Fresno PM2.5
- The highest PM mean is in Node 2 -- 85 ug/m**3
- Predicted by low minimum temps and moderate stability
- The 85 ug/m**3 is three time higher than other nodes but with only 62 occurrences compared to 635 total PM2.5 records analyzed

Figure 2 – Tree for Fresno PM2.5



Summary of All Seasons, Fall, and Winter Analyses

- Table 1 shows the analyses for all seasons, in all of the four regions
  - Stability is the most important met variable followed by min temp, and visibility
- Tables 2 and 3 show Winter and Fall analyses
  - -In winter, there is a dependence on both stability and mintemp with one node having visibility as a primary splitter
  - - There appears to be no dependency on wind speeds in Corcoran case for Fall

Table 1 -- All Seasons Analysis

Site	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	No Others	93	41	Stab - 100
Bakersfield	PM25	Visibility	Stability	71	14	Vis – 100, Stab –20, mintemp – 50
Fresno	PM10	Stability	Mintemp, Wind speed, and RH	95	14	Stab –100, wind speed – 82, mintemp – 66, and RH - <25
Fresno	PM25	Mintemp	Stability, Visibility	84	33	Mintemp –100, stab – 74, vis – 39
Stockton	PM10	Stability	No Others	63	<30	Stab – 100, vis – 11
Stockton	PM25	Mintemp	Visibility	45	<25	Mintemp –100, vis – 44
Corcoran	PM10	Stability	Maxtemp	96	21	Stab – 100, maxtemp – 53
Corcoran	Coarse (PM10-PM25)	Maxtemp	Stability	84	10	Stab – 100, Meantemp – 81, Maxtemp – 79

Table 2 -- Winter Analysis

Site	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	Mintemp, RH	112	22	Stab –100, mintemp 56, RH – 29
Bakersfield	PM25	Visibility	Mintemp	86	8	Vis – 100, Stab –97, mintemp – 70
Fresno	PM10	Mintemp	Stability	236	34	Mintemp –100, stab - 36
Fresno	PM25	Stability	Mintemp, Visibility	127	16	Stability –100, mintemp – 89, vis – 39
Stockton	PM10	Mintemp	No Others	75	41	Mintemp – 100
Stockton	PM25	Mintemp	No Others	52`	21`	Mintemp –100

Table 3 -- Fall Analysis

Site	PM Pollutant Type	Primary Splitting Variable	Secondary Splitting Variables	Highest Mean PM Concentration (ug/m**3)	Lowest Mean PM Concentration (ug/m**3)	Variable Importance
Bakersfield	PM10	Stability	No Others	93	41	Stab -100
Bakersfield	PM25	Visibility	Stability	71	14	Vis – 100, Stab –20, mintemp – 50 (did not play a role)
Fresno	PM10	Stability	Mintemp, Wind speed, and RH	95	14	Stab –100, wind speed – 82, mintemp – 66, and RH - < 25
Fresno	PM25	Mintemp	Stability, Visibility	84	33	Mintemp –100, stab – 74, vis – 39
Stockton	PM10	Stability	No Others	63	<30	Stab – 100, vis – 11
Stockton	PM25	Mintemp	Visibility	45	<25	Mintemp –100, vis – 44
Corcoran	PM10	Stability	Maxtemp	96	21	Stab – 100, maxtemp – 53
Corcoran	Coarse (PM10-PM25)	Maxtemp	Stability	84	10	Stab – 100, Meantemp – 81, Maxtemp – 79

CONCLUSIONS

- All Season Analysis
  - PM2.5 – more complex trees with mintemp and visibility being the prime splitters
  - PM10 -- Stability is the primary variable for all four sites
  - Stability greater than 8 (850mb Temp-sfc min temp) defines in most cases very high mean PM10 concentrations, considering other variables
  - Low visibility, low mintemps, and high stability seem to predict high mean PM2.5 concentrations

CONCLUSIONS (Part II)

- Winter Runs (December, January, and February)
  - Winter season splitting variables for PM10 are not so stability dependent – other variables include mintemp at two sites. Secondary splitters include mintemp, stability, and relative humidity
  - PM2.5 – primary splitters were visibility, stability, and mintemp. Secondary splitters include mintemp and visibility
  - The highest PM10 occurred when stability > 12 (very high) and RH < 82 or in another node when moderate stability and low overnight temps. High PM10 levels were coincident with high PM2.5 levels

CONCLUSIONS (Part III)

- Fall Runs (September, October, November)
  - Difficult to tell when fall ends and winter begins so November end time is arbitrary
  - Stability was primary PM10 splitter for all 3 locations. Other secondary splitters were visibility and RH
  - For PM2.5, primary splitter was visibility at all 3 sites. Other secondary variables included meantemp (compared to mintemp for winter), stability, and RH

KEY HIGHLIGHTS OF CART

- CART was used successfully to categorize what met variables produced the highest and lowest PM levels in the San Joaquin Valley and supported the SJV PM10 SIP
- CART simulated three regions of the valley quite well
- CART worked well with regional scale meteorology (over all or part of the Valley) but did not simulate localized high wind speeds over the Corcoran area